METHODS AND SYSTEMS FOR

CALCULATING REQUIRED SCRATCH MEDIA

Background of the Invention

[0001]   Backup applications use scratch media to process backup jobs that backup data. Scratch media are either new media or media that contain data that can be overwritten. In order for the backup jobs to execute properly, scratch media of the appropriate type must be loaded into the backup devices, such as tape libraries, used by the backup applications.

[0002]   Currently, a user must manually estimate how much scratch media are needed for future backup jobs. If the user does not correctly estimate how much scratch media are needed before the backup jobs run, the backup jobs halt while awaiting additional scratch media to be loaded. The backup job may fail if the media are not loaded within a predetermined time period.

[0003]   The manual estimating process may be highly complex. Over a given time period, hundreds of backup jobs may execute using backup devices located in multiple data centers. Each backup job may be a full or incremental backup of clients with varying data sizes. Additionally, the estimating process is

further complicated by the use of media pools by the backup applications. Each backup job may be associated with one or more media pools and may not be able to access scratch media in other media pools. Therefore, the user must not only ensure that each backup device has the appropriate amount of scratch media loaded but also must ensure that each media pool contains the appropriate amount of scratch media to execute the backup jobs using that media pool. Because the estimating process is so complex, backup jobs often fail because the appropriate amount of scratch media has not been loaded.

Summary of the Invention

[0004]   Methods and systems are disclosed for calculating required scratch media. In one embodiment, backup job information is obtained for a plurality of backup jobs. The backup job information is obtained from one or more backup applications and is used to calculate a required number of scratch media needed for one or more future executions of at least one of the backup jobs. The required number of scratch media is then presented to a user.

Brief Description of the Drawings

[0005]   Illustrative embodiments of the invention are illustrated in the drawings in which:

[0006]   FIG. 1 illustrates an exemplary configuration of a media manager to manage scratch media for one or more backup applications;

[0007]   FIG. 2 is a flow diagram illustrating an exemplary method for managing scratch media that may be used by the media manager of FIG. 1;

[0008]   FIG. 3 illustrates an exemplary presentation to a user of a required number of scratch media;

[0009]   FIG. 4 illustrates a second exemplary presentation to a user of a required number of scratch media;

[0010]   FIG. 5 is a flow diagram illustrating an exemplary method for calculating required scratch media that may be used by the method of FIG. 2;

[0011]   FIG. 6 is a flow diagram illustrating an exemplary method for confirming media to be used for the required scratch media; and

[0012]   FIG. 7 illustrates an exemplary user interface that may be used to enter media to be used for the required scratch media.

Detailed Description

[0013] An exemplary configuration using a media manager to manage scratch media is illustrated in FIG. 1. It should be understood that that the term media should not be regarded as being limited to two or more pieces of media, but could also include a single medium. The media manager 100 may be responsible for managing scratch media for multiple backup applications 110, 120. In one embodiment, backup application 110 may be Hewlett Packard Corporation's (a Delaware corporation having its principle place of business in Santa Clara, CA) Data Protector. Other backup applications may also be managed by media manager 100.

[0014] Backup application 110 is responsible for backup jobs 112, 114. Similarly, backup application 120 is responsible for backup jobs 122, 124. The backup jobs 112, 114, 122, 124 may each be full or incremental backups of data contained in various locations, such as on backup server(s) executing the backup applications 110, 120, one or more clients communicatively coupled to the backup servers, or other location (e.g., disk device) storing data to be backed up. It should be appreciated that in alternate embodiments, each backup application 110, 120 may be responsible for a larger number of backup jobs than that illustrated in FIG. 1. Additionally, media manager 100 may be responsible for managing the scratch media for a different number of backup applications 110, 120 than that shown.

[0015] Each backup application 110, 120 may be communicatively coupled to one or more backup devices (not shown) to backup the data

associated with a backup job 112, 114, 122, 124. By way of example, backup devices may be standalone tape drives, robotic tape libraries, or optical disk libraries. The backup devices may each use scratch media to backup the data associated with the backup jobs 112, 114, 122, 124. Scratch media may be new media or previously written media that may now be overwritten. In one embodiment, the scratch media may be distributed into one or more media pools, each media pool associated with one or more of the backup jobs. Each of the media pools may contain media located in one backup device or multiple interconnected backup devices (e.g., interconnected tape libraries). In this embodiment, the backup jobs may only use scratch media contained in their associated media pools.

[0016]   Backup applications 110, 120 are communicatively coupled to media manager 100. Media manager may obtain backup job information from backup applications 110, 120 for their respective backup jobs. By way of example, backup applications may be communicatively coupled to media manager via an internet or intranet and may send backup job information to the media manager using an open communication, such as XML (Extensible Markup Language). The backup job information may include devices and/or media pools used by the backup jobs, scratch media currently available in each of the backup devices and/or media pools, types of media used, compression rates, types of the backup jobs (e.g., full or incremental), locations of the data being backed up by the backup jobs, historical backup size information for the backup jobs, future scheduled executions of the backup jobs, and any other information that may be

used to estimate the required number of media needed for future executions of the backup jobs.

[0017]   Media manager 100 may include an integration agent 102 to request and/or receive the backup job information from the backup applications 110, 120. Additionally, media manager 100 includes planner 104 that uses the backup job information to calculate a required number of scratch media needed by the backup applications 110, 120 to execute one or more future backup jobs. Media manager may also include a user interface 106 to present the required number of scratch media to a user. The calculation of the required number of scratch media and the presentation to a user will be described in further detail below.

[0018]   Media manager 100, integration agent 102, planner 104, and user interface 106 may be implemented in software, firmware, hardware, or a combination of these. Additionally, integration agent 102, planner 104, and user interface 106 may reside on separate physical locations, such as separate servers. By way of example, user interface 106 may be a web browser on a remote client.

[0019]   It should be appreciated that configurations different than the one shown in FIG. 1 may be used in alternate embodiments. For example, integration agent 102 may additionally or alternately reside on one or more of the servers executing the backup applications managed by the media manager 100. This configuration may facilitate the exchange of information in configurations containing firewalls.

[0020]   FIG. 2 illustrates an exemplary method of managing scratch media that may be used by media manager 100. Backup job information is obtained 205 from one or more backup applications 110, 120 for a plurality of backup jobs 112, 114, 122, 124. As previously described, backup job information may include backup devices and/or media pools used by the backup jobs, scratch media available in each of the devices and/or media pools, types of media used, compression rates, types of the backup jobs (e.g., full or incremental), locations of the data being backed up by the backup jobs, historical backup size information for the backup jobs, future scheduled executions of the backup jobs, and other information that may be used to estimate a required number of scratch media needed for future executions of the backup jobs.

[0021]   Next, the required number of scratch media needed for one or more future executions of at least one of the backup jobs is calculated 210 using the backup job information. Planner 104 may estimate the required number of scratch media required for the future executions of backup jobs scheduled within a predetermined period of time. It should be appreciated that the calculation may be performed in any manner that estimates the number of scratch media that will be required for future executions of one or more of the backup jobs.

[0022]   In one embodiment, a user may divide the backup jobs into one or more subsets and may configure media manager 100 to perform the calculation 210 for one or more of these subsets at different intervals of time. By way of example, a user may configure media manager 100 to perform the calculation for subset A of the backup jobs on a daily basis and to perform the calculation for

subset B of the backup jobs on Tuesdays and Thursdays. Backup jobs may be divided into subsets according to data center, importance of data being backed up, or other criteria determined by the user. The number of scratch media that the backup jobs in a subset is estimated to require may be calculated 210 for the time period before the next calculation for that subset is scheduled to be performed.

[0023]   The calculated required number of scratch media is then presented 215 to a user. FIGS. 3 and 4 illustrate two exemplary embodiments used to present the information to a user. As shown in FIG. 3, user interface 106 may present the information to a user on a screen 300 containing a column 310 for each device used by the future executions of the backup jobs for which the calculation 210 was performed. A second column 320 may contain the required number of scratch media needed for each of the devices listed in column 310.

[0024]   Alternately, or in addition to, the presentation illustrated in FIG. 3, the required number of scratch media may also be presented to the user for each media pool. Screen 400 contains a column 410 for each media pool used by the future executions of the backup jobs for which the calculation 210 was performed. The number of media required for each of the media pools is listed in column 420.

[0025]   It should be appreciated that the required number of scratch media may be presented to a user in a different manner than that illustrated in FIGS. 3 and 4. Additionally, in alternate embodiments, the user may be presented with other information that may facilitate the management of scratch media. By way

of example, the user may be presented with a report identifying media pools having a greater amount of scratch media than the calculated 210 required scratch media. As a second example, the user may be presented with a report similar to that shown in FIG. 3 when the number of scratch media available in the pool becomes lower than a default value or a estimated number of media required based on historical usage of scratch media in the media pool

[0026]   FIG. 5 illustrates an exemplary method that may be used to calculate 210 the required number of scratch media needed for one or more future executions of at least one backup job. The method begins by obtaining 505, for a future execution of a backup job, the type of the backup job to be performed. By way of example, the backup type may be a full or incremental backup. In alternate embodiments, the backup type may not need to be obtained 505 as each backup job may be associated with a predetermined type of backup.

[0027]   Next, an average historical backup size for the type of backup to be executed is obtained 510. The average historical size may be calculated by planner 104 by averaging backup sizes for previous executions of the backup job. Alternately, the average historical backup size may be included in the backup job information obtained from the backup application responsible for the backup job.

[0028]   The number of media required for the future execution is then estimated at 515 by dividing the average historical backup size by the average capacity of the media type used by the backup job. In some embodiments (e.g., embodiments in which the media is defined as non-appendable), the required

number of media may be rounded to the next or nearest whole number. In other embodiments (e.g., where the media may be appended), the required number of media may be rounded after totaling the number of media required for all backup jobs and subtracting from an available existing capacity (e.g., at 535). If the average backup size is not available, either because the backup job has not previously executed or other reason, a default number of scratch media, that may be based on the type of backup device and/or the type of backup job, may be used to estimate the required number of scratch media needed for the future execution of the backup job. Planner 104 may calculate the average capacity of the media type by obtaining, from the backup job information, the type of media used for the backup job and any compression information for the backup job or media pool used by the backup job. Alternately, the backup job information may include the average capacity of the media type used by the backup job.

[0029] If there are more future executions of additional backup jobs 520 which are included in the calculation 210, the method continues back at 505 to obtain the backup type for another future execution. It should be appreciated that if the calculation 210 includes multiple future executions of the same type of the same backup job, the number of media may be calculated for those executions by multiplying the number of executions by the amount obtained from 515.

[0030] If there are no more future executions of backup jobs to process 520, the method may continue by totaling 525 the number of media required for each media pool. The existing number of scratch media that will be available in

each media pool is also determined 530. The existing number of scratch media available in each pool may be determined from information included in the backup job information.

[0031]   Additional factors may be taken into account when determining the existing number of scratch media. By way of example, a determination may be made if a protected period for one or more existing data media in the media pool has expired, which may mean the data media may be overwritten and used as scratch media. A second factor that may be taken into account is the availability of appendable media (media that contains data that may be appended by a future execution of a backup job). The number of media required for each media pool may be totaled 525 for both non-appendable media (for backup jobs that may not append to existing media) and appendable media (for backup job that may append to existing media).

[0032]   For each media pool, the total number required is subtracted 535 from the existing number of scratch media contained in the pool. In embodiments using appendable media, the existing capacity may also be subtracted from the total capacity required for appendable media. The remaining required capacity may then be divided by an average capacity of the media type to obtain a required number of appendable media.   This total may then be combined with the estimated required total number of non-appendable media.

[0033]   In an alternate embodiment, in addition or instead of 525-535, a total number of media may be calculated for the backup devices used by the future backup jobs. The existing number of scratch media available in the

backup devices is also determined. The estimated total number of required media may then be subtracted from the existing total available in the backup device to obtain a number of media required for each backup device. When determining the number of existing scratch media available in the backup devices, appendable media may be taken into account as described above.

[0034]   It should be appreciated that the order of the method depicted in FIG. 5 may be changed. For example, the determination of the number of existing scratch media 530 may be made prior to calculating the total media required for each media pool 525. Additionally, it should be appreciated that in alternate embodiments, an estimated number of media required for a media pool may be calculated and presented to a user in a different manner than that described with reference to FIGS. 2 and 5. For example, historical usage of scratch media in a media pool may be analyzed and the user may be prompted to initialize new scratch media for the media pool when the available number of scratch media becomes too low.

[0035]   FIG. 6 illustrates an exemplary method for confirming media to be used for the required scratch media that may be used by media manager 100. After presenting the required number of scratch media to a user 215, media manager 100 may receive 605 a list from the user of one or more media to be used for at least a portion of the required scratch media. For each media in the list, planner 104, or another component of media manager 100, may then determine 610 if the media that is intended to be used for the required scratch media is valid scratch media.

[0036] In order to determine if a media is valid scratch media, various factors may be taken into account. One factor that may be used is whether a protection date has expired for a previously written data media. Other factors such as the quality of the media (which may be determined from backup job information based on previous problems encountered by backup applications 110, 120 in using the media) or whether the media is included or initialized in the appropriate media pool may also be used to determine 610 if the media may be used to meet the scratch media requirements for the future executions of one or more backup jobs.

[0037] A screen 700 that may provided by user interface 106 for the user to enter a list of one or more media to be used for the required scratch media is illustrated in FIG. 7. A text-entry area 702 is provided for the user to enter a media pool for which the media is to be entered. In alternate embodiments, the user interface may provide a list of media pools requiring scratch media from which the user may select, or the screen 700 may provide areas for the user to enter media for a backup device or combination of backup device and media pool.

[0038] A display area 704 displays the number of media required for the media pool. It should be appreciated that in alternate embodiments, screen 700 may not include display area 704 or may include additional or alternate display areas to display the number of media required for a selected backup device.

[0039] Screen 700 further includes a text entry area 706 to enter the media to be used for the scratch media for the device and/or media pool. The

user interface 106 may receive this information, for example, from a keyboard or a barcode scanner used to scan barcodes located on the media. After the media is entered, a verify button 708 is provided to determine 610 if the media are valid scratch media. Alternately, the user may be prompted to verify or skip verification as each piece of media is entered.

[0040]   A column 710 listing the media entered by the user may also be included in screen 700. The device for which the media is to be entered is listed in a second column 720. As the media list may be lengthy, scroll bar 730 may also be provided for the user to scroll through the list. In alternate embodiments, the second column 720 may display the media pool for which the media is to be entered or an additional column may be provided to display this information.

[0041]   It should be appreciated that the methods described above may be performed by hardware components or may be embodied in sequences of machine-executable instructions, which may be used to cause a machine, such as a general-purpose or special-purpose processor or logic circuits programmed with the instructions to perform the actions set forth in FIG. 2, 5, and 6. Alternatively, the methods may be performed by a combination of hardware and software.